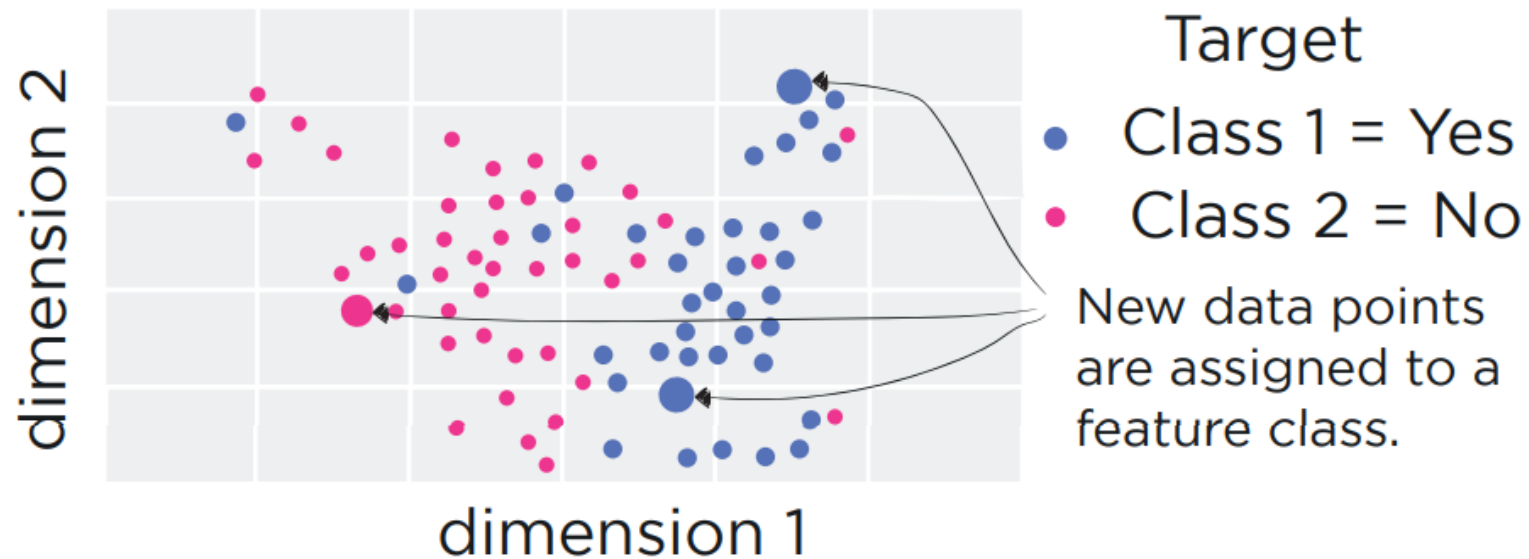


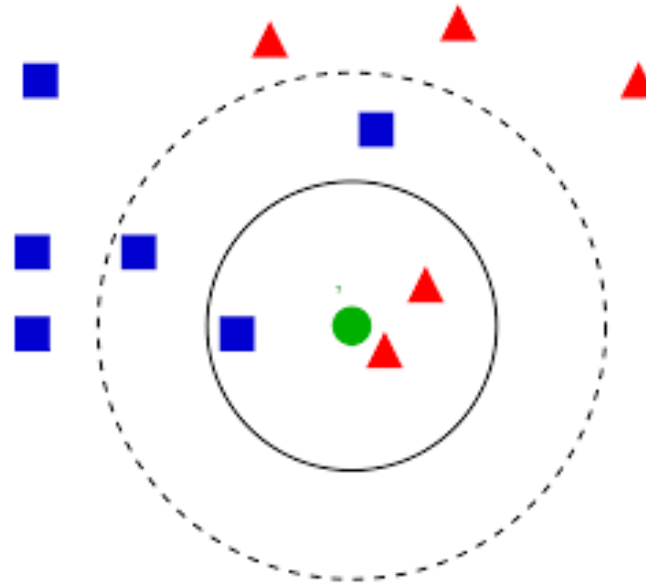
# CT-562 MACHINE LEARNING

NED University of Engineering & Technology

# INSTANCE-BASED LEARNING

Instance-based learning are the systems that learn the training examples by heart and then generalizes to new instances based on some similarity measure. It is called instance-based because it builds the hypotheses from the training instances. It is also known as memory-based learning or lazy-learning (because they delay processing until a new instance must be classified).





# K-NEAREST NEIGHBOR(KNN) ALGORITHM

## OVERVIEW

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning techniques.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

# KNN

- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

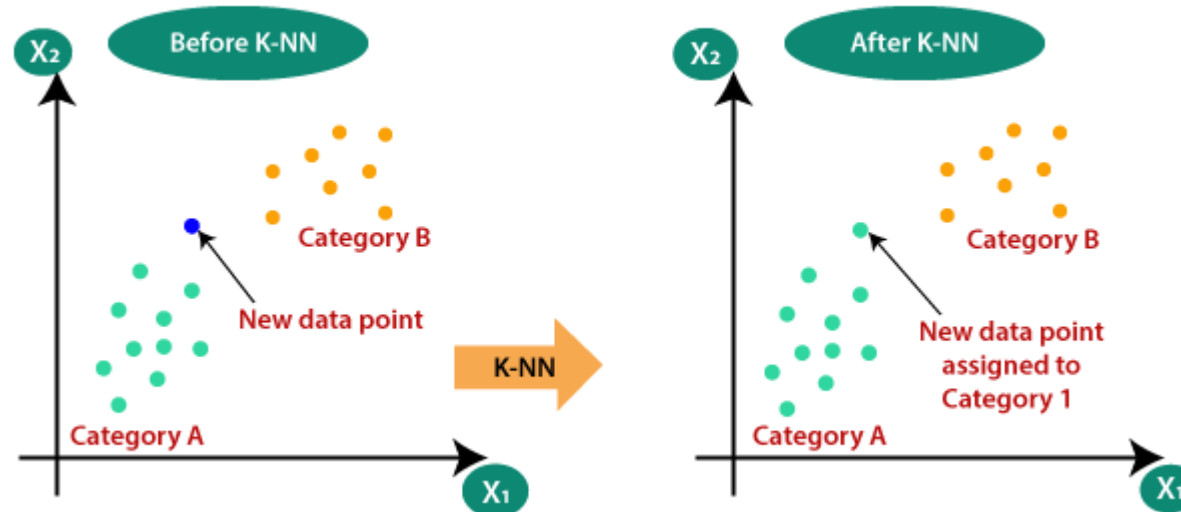
## SIMPLE EXAMPLE

- Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.



## WHY DO WE NEED A K-NN ALGORITHM?

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point  $x_1$ , so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:



# HOW DOES K-NN WORK?

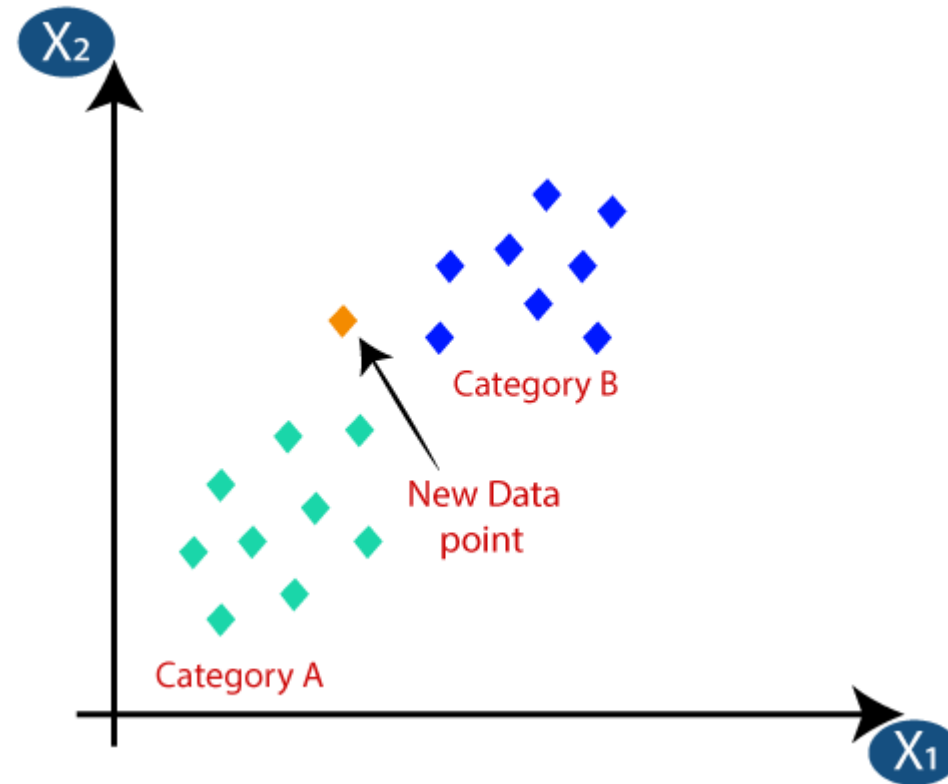
The K-NN working can be explained on the basis of the below algorithm:

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Among these k neighbors, count the number of the data points in each category.
- Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
- Step-6: Our model is ready.



# HOW DOES K-NN WORK?

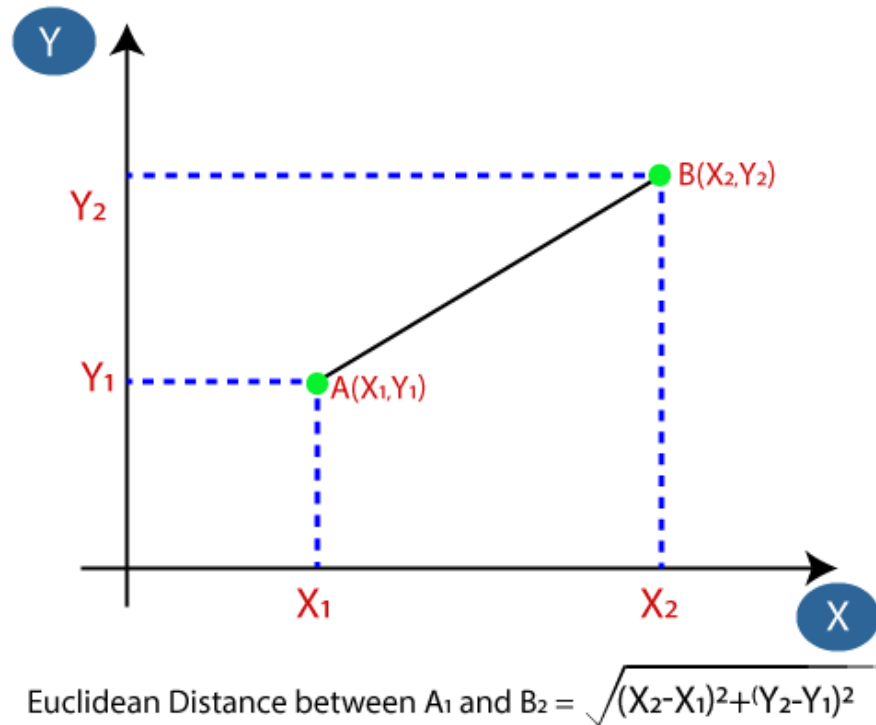
Suppose we have a new data point and we need to put it in the required category. Consider the below image:



# HOW DOES K-NN WORK?

- Firstly, we will choose the number of neighbors, so we will choose the  $k=5$ .
- Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:

Euclidian distance: squareroot of sum of squares of differences for two features:  $(\Delta x)^2 + (\Delta y)^2$



## HOW DOES K-NN WORK?

- By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:



As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

## EXAMPLE

Name	Age	Gender	Sport
Ahmed	32	M	Football
Mark	40	M	Neither
Sara	16	F	Cricket
Zaira	34	F	Cricket
Saleem	55	M	Neither
Rahul	40	M	Cricket
Saba	20	F	Neither
Smith	15	M	Cricket
Nazish	55	F	Football
Michael	15	M	Football
Mirha	5	F	??

## EXAMPLE

Name	Age	Gender	Sport
Ahmed	32	0	Football
Mark	40	0	Neither
Sara	16	1	Cricket
Zaira	34	1	Cricket
Saleem	55	0	Neither
Rahul	40	0	Cricket
Saba	20	1	Neither
Smith	15	0	Cricket
Nazish	55	1	Football
Michael	15	0	Football
Mirha	5	1	??

K = 3

## EXAMPLE

Euclidian distance: squareroot of sum of squares of differences

for two features:  $(\Delta x)^2 + (\Delta y)^2$

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

## EXAMPLE

Name	Age	Gender	Distance	Sport
Ahmed	32	0	27.02	Football
Mark	40	0	35.01	Neither
Sara	16	1	11.00	Cricket
Zaira	34	1	9.00	Cricket
Saleem	55	0	50.01	Neither
Rahul	40	0	35.01	Cricket
Saba	20	1	15.00	Neither
Smith	15	0	10.00	Cricket
Nazish	55	1	50.00	Football
Michael	15	0	10.05	Football
Mirha	5	1		??



THANK YOU